УДК 004.891:004.827:004.931

А.Ю. Максимова
Институт прикладной математики и механики НАН Украины, Украина
Украина, 83050, г. Донецк, ул. Розы Люксембург, 74

Метод принятия решения для модели классификации, построенной на основе внутриклассовой кластеризации FCM-алгоритмом

A. Maksymova
Insitute of Applied Mathematics and Mechanics NAS of Ukraine, Ukraine
Ukraine, 83050, c. Donetsk, Roza Luksemburg av., 74

Decision Making Method for Classifying Models Based on Intra-class Clustering on FCM-algorithm

О.Ю. Максимова
Інститут прикладної математики і механіки НАН України, Україна
Україна, 83050, м. Донецьк, вул. Рози Люксембург, 74

Метод прийняття рішень для моделі класифікації, що заснована на внутрікласової кластеризації FCM-алгоритмом

В статье предложен метод распознавания образов для задач с сильно пересекающимися классами образов. В рамках метода реализован алгоритм поиска оптимального в соответствии с некоторым критерием адекватности числа кластеров FCM-алгоритма и построена модель, описывающая структуру данных задачи распознавания, а также предложен метод принятия решения по этой модели.
Ключевые слова: нечеткий классификатор, нечеткая кластеризация, принятие решений.

In this paper we propose a method for pattern recognition problems with strongly overlapping classes of images. The method for optimal according to some validation measure number of clusters for FCM-algorithm is implemented. The classification model that describes the data structure for pattern recognition problem and the decision making method for this model are proposed.
Key words: fuzzy classifier, fuzzy clustering, decision making.

У статті запропоновано метод розпізнавання образів для задач з класами образів, що сильно перетинаються. У рамках методу реалізовано алгоритм пошуку оптимального відповідно з деяким критерієм адекватності числа кластерів FCM-алгоритму і побудовано модель, що описує структуру даних задачі розпізнавання, а також запропоновано метод прийняття рішення за цією моделлю.
Ключові слова: нечіткий класифікатор, нечітка кластеризація, прийняття рішень.

Введение

При построении классификационных моделей для решения задач распознавания образов используют методы машинного обучения, например, машину опорных векторов, нейронные сети и т.д. Однако высокая точность таких моделей достигается за счет их сложности и, как следствие, низкой прозрачности. Все методы распознавания разделя-

ют на две группы: дискриминативные алгоритмы, которые строят разделяющую границу между классами; и генеративные алгоритмы, которые строят прототипы классов образов. Классификационные модели, построенные в рамках дискриминативного подхода для задач распознавания с существенно пересекающимися классами образов, обладают низкой обобщающей способностью [1]. С другой стороны, генеративный подход позволяет решить эту проблему. Однако, четкие методы в спорных ситуациях на границах классов всегда будут давать некоторую ошибку. В таком случае проблема может быть решена с использованием нечеткого подхода, когда результатом работы алгоритма будет вектор степеней уверенности, с которой объект отнесен к тому или иному классу образов. Тогда для объектов, попадающих в области пересечения, т.е. типичных для нескольких классов образов, будет получена существенная информация, на основании которой можно принять корректное решение. Методы распознавания образов, в которых используются нечеткие множества, объединяют под названием нечеткие классификаторы [2]. Существуют несколько подходов для построения нечетких классификаторов: настройка базы ЕСЛИ-ТО правил эволюционными алгоритмами [3], [4], применение FyzzyLVQ и FSOM сетей [5], использование методов нечеткой кластеризации [6].

Предложенная Л. Заде концепция информационных гранул, которая заключается в разбиении множества исходных данных на элементарные составляющие, может быть применена в рамках генеративного подхода к распознаванию. Одним из способов представления совокупности информационных гранул являются кластеры, а задача построения информационных гранул в условиях неопределенности решается методами нечеткой кластеризации [7, с. 203].

Цель работы: В данной статье предлагается новый способ построения нечеткого классификатора. В первой части статьи приведена общая постановка задачи распознавания образов в условиях неопределенности. Во второй части описан метод нечеткой кластеризации, основанный на известном алгоритме нечетких *c*-средних (FCM-алгоритме), который позволяет решить проблему поиска оптимального количества кластеров с точки зрения определенного критерия адекватности нечеткого разбиения. В третьей части описан новый метод принятия решения о принадлежности объекта определенному классу образов. В четвертой части на основании экспериментов показана эффективность предложенного метода.

1 Общая постановка задачи распознавания образов в условиях неопределенности

Далее везде под R, $\mathbf{N}$, 2^M будем, как обычно, понимать соответственно множество действительных чисел, множество натуральных чисел, булеан множества M.

Пусть O – множество объектов какой-либо природы, например, бензинов или вин. Пусть $\{f_i\}_{i\in\{1,\dots,m\}}$ представляет собой семейство отображений $f_i : O \to X_i$, ставящих в соответствие каждому объекту из O значение i-го признака из множества X_i. Область значений X_i i-го признака может быть, например, подмножеством вещественных чисел. Пусть $\mathsf{A} \subseteq 2^O$ – множество классов объектов. Пусть Ω – некоторое множество. Поставим во взаимооднозначное соответствие каждому элементу $A \in \mathsf{A}$ элемент множества Ω – имя класса объектов. Множество Ω назовем алфавитом классов объектов.

Тогда алгоритм распознавания формально может быть представлен функцией $D: X_1 \times \ldots \times X_m \to \widetilde{\Omega}$, где $\widetilde{\Omega}$ – пространство нечетких множеств над Ω.

Мы будем рассматривать задачу построения модели распознавания, реализующей функцию D по выборке прецедентов, заданной в виде конечного множества $Z \subseteq X_1 \times X_2 \times \ldots \times X_m \times \Omega$. На основании построенной модели будем строить систему распознавания образов.

Для открытых систем распознавания с течением времени могут появляться новые объекты с известными принадлежностями классам образов, которые должна учитывать модель. В связи с этим введем параметр t – дискретное время, от которого будет завистить обучающее множество $Z_t = Z(t)$. Выборку Z в начальный момент времени назовем Z_0,а построенный по ней алгоритм распознавания D_0. Необходимо разработать алгоритм дообучения классификационной модели.

2 Алгоритм кластеризации с неизвестным числом кластеров

В данной работе для решения задачи распознавания предлагается выполнять предварительный анализ данных с целью установления внутриклассовой структуры для каждого класса образов. Для этих целей предлагается использовать FCM-алгоритм нечеткой кластеризации. Данный алгоритм для множества прецедентов $X = \{\mathbf{x}_i\}_{i=1}^{n}$ строит нечеткое c-разбиение в виде матрицы $U = [u_{ik}]_{n \times c}$, где u_{ik} – степень принадлежности элемента $\mathbf{x}_i$ кластеру с номером k, а c – количество кластеров, которая является параметром алгоритма. Рассмотрим три типа c-разбиений, которые будут использованы далее в работе:

$$M_{pcn} = \{U \in \mathrm{R}^{c \times m} \mid \forall i, k\ 0 \le u_{ik} \le 1;; \forall k\ \exists i\ u_{ik} > 0;; \forall i\ 0 < \textstyle\sum_{i=1}^{n} u_{ik} \le n\ \}, \tag{1}$$

$$M_{fcn} = \left\{ U \in M_{pcn} \middle| \forall k \textstyle\sum_{i=1}^{c} u_{ik} = 1 \right\}, \tag{2}$$

$$M_{hcn} = \left\{ U \in M_{fcn} \middle| \ \forall k, i\ u_{ik} \in \{0,1\} \right\}, \tag{3}$$

где M_{pcn} – возможностное, M_{fcn} – нечеткое и M_{hcn} – четкое c-разбиения.

Кроме разбиения $U \in M_{fcn}$, результатом работы FCM-алгоритма являются геометрические центры кластеров $\mathbf{G} = \{\mathbf{g}_i, \mathbf{g}_2, \ldots, \mathbf{g}_c\} \subset \mathrm{R}^m$. FCM-алгоритм минимизирует функционал Бездека-Данна:

$$J_{\gamma}^{FCM}(U, \mathbf{G};; X) = \textstyle\sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^{\gamma} d^2(\mathbf{x}_k, \mathbf{g}_i) \to \underbrace{min}_{\{U, \mathbf{G}\}} \tag{4}$$

при ограничениях:

$$\textstyle\sum_{i=1}^{c} u_{ik} = 1, \forall x_k, k = \overline{1, n}, \tag{5}$$

где γ – коэффициент нечеткости, а $d^2(\mathbf{x}, \mathbf{g})$ – квадрат расстояния между элементом $\mathbf{x}$ и центром кластера $\mathbf{g}$. В данной работе рассматривается евклидово расстояние.

FCM-алгоритм ищет глобальный оптимум методом итеративной оптимизации, на каждом следующем шаге вычисляя по центрам кластеров степени принадлежности u_{ik} для объекта $\mathbf{x}_k$:

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{d(\mathbf{x}_k, \mathbf{g}_i)}{d(\mathbf{x}_k, \mathbf{g}_j)} \right)^{\frac{2}{\gamma-1}} \right)^{-1}, 1 \le i \le c, 1 \le k \le n \qquad (6)$$

а затем новые центры кластеров по u_{ik} :

$$\mathbf{g}_k = \frac{\sum_{\mathbf{k=1}}^{\mathbf{n}} (u_{ik}^{\gamma} \cdot \mathbf{x}_k)}{\sum_{\mathbf{k=1}}^{\mathbf{n}} u_{ik}^{\gamma}}. \qquad (7)$$

Инициализация алгоритма осуществляется за счет определения начальных значений прототипов кластеров. По обучающей выборке Z вычисляются минимальные p_{1f_i} и максимальные p_{2f_i} значения для каждого информативного признака f_i, $i = \overline{1,m}$, а координаты центров кластеров вычисляются по формуле:

$$\mathbf{g_k} = \mathbf{p}_1 + \frac{k(\mathbf{p_2} - \mathbf{p_1})}{(c+1)}, \qquad (8)$$

где $\mathbf{p}_1 = (p_{1f_1}, p_{1f_2}, \dots, p_{1f_m})$, а $\mathbf{p}_2 = (p_{2f_1}, p_{2f_2}, \dots, p_{2f_m})$, $1 \le k \le c$, c – количество кластеров.

Основным недостатком FCM-алгоритма является необходимость задавать в качестве параметра алгоритма количество кластеров c, что при исследовании структуры данных заранее неизвестно. Для решения проблемы поиска оптимального числа кластеров используют критерии оценки адекватности кластеров, что неформально можно описать как «наиболее подходящая кластерная структура», которая для каждой задачи будет отличаться, поэтому и выбор критерия будет отличаться.

Далее рассмотрим некоторые из возможных критериев адекватности кластеров. В четвертой части статьи будут рассмотрены экспериментальные результаты применения данных критериев для некоторых задач распознавания в рамках предложенной в данной работе модели классификации.

Коэффициент разбиения Бездека [8]:

$$v_{PC}(U) = \frac{\sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^2}{n} \qquad (9)$$

где U – нечеткое c-разбиение, c –количество кластеров, n – количество элементов обучающей выборки X . Для v_{PC} известны следующие свойства:

$$v_{PC} = 1 \Leftrightarrow U \in M_{hcn}, \qquad (10)$$

$$v_{PC} = \frac{1}{c} \Leftrightarrow U = \left[\frac{1}{c} \right] = \overline{\mathrm{U}}, \qquad (11)$$

где M_{hcn} определено в (3), $\overline{U}$ – «самое нечеткое» разбиение, так как оно каждый элемента из X с одинаковой степенью уверенности относит к каждому из c-кластеров. Критерий разбиения Бездека относится к классу критериев, которые используют только информацию о разбиении U, но не использует информацию о самих данных, такую как прототипы кластеров $\mathbf{G}$ и X .

К классу критериев адекватности, которые используют полную информацию $(U,\mathbf{G};X)$, относится индекс Кси-Бени:

$$v_{XB}(U,\mathbf{G};X)=\frac{\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{2}\|\mathbf{x}_k-\mathbf{g}_i\|^2}{n\,(\underbrace{\min}_{i\neq j}\{\|\mathbf{g}_i-\mathbf{g}_j\|^2\})}=\left[\frac{\left(\frac{\sigma}{n}\right)}{sep(\mathbf{G})}\right], \quad (12)$$

где σ – полная вариация по $(U,\mathbf{G})$, а $sep(\mathbf{G})$ – разделение по $\mathbf{G}$

$$\sigma(U,\mathbf{G};;X)=\sum_{i=1}^{c}\left(\sum_{k=1}^{n}u_{ik}^{2}\|\mathbf{x}_k-\mathbf{g}_i\|^2\right), \quad (13)$$

$$sep(\mathbf{G})=\underbrace{\min}_{i\neq j}\left\{\|\mathbf{g}_i-\mathbf{g}_j\|^2\right\}. \quad (14)$$

Чем меньше значение v_{XB}, тем лучшее разбиение на X получено, что верно при $\gamma=2$ [8]. Исследования влияния коэффициента нечеткости на значение индекса Кси-Бени (12) говорят о его нестабильности для больших значений γ [8].

Предлагается искать оптимальное с точки зрения критерия v^* количество кластеров по следующей схеме (рис. 1). Результатом работы предложенного метода будет модель $FP^{\omega}=<U,\mathbf{G},X,c>$, которая, в рамках общей, предлагаемой автором, концепции, будет называться нечетким портретом класса образов $\omega\in\Omega$, где Ω – алфавит классов образов задачи распознавания [9].

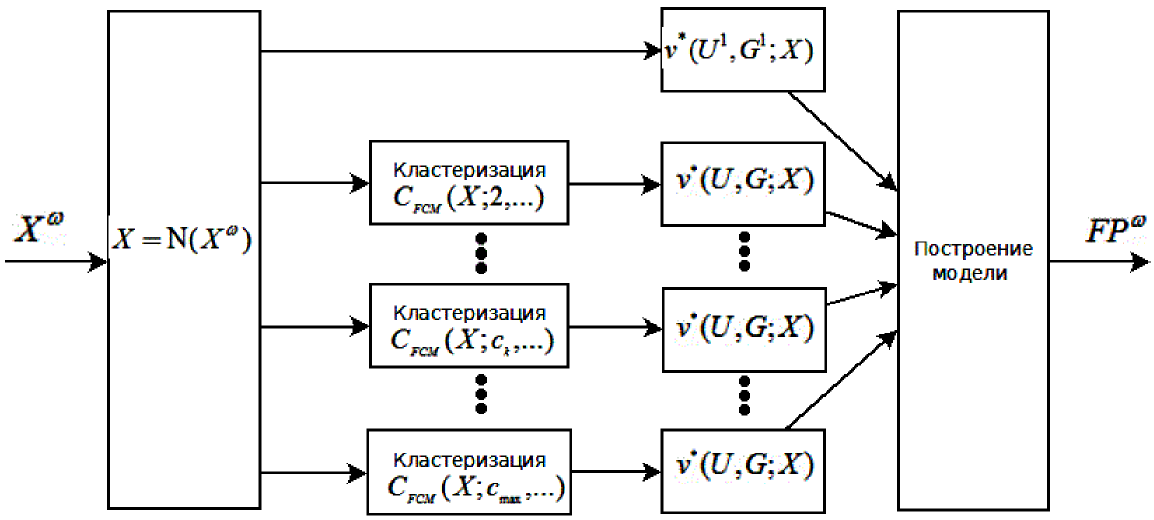


Рисунок 1 – Общая схема построения модели класса образов

На схеме $X^{\omega}\subset\mathrm{R}^m$ соответствует входным данным алгоритма, причем индекс ω указывает на то, что схема решает задачу внутриклассового анализа данных, где $\omega\in\Omega$ – имя анализируемого класса образов, однако в общем случае это не обязательно; $\mathrm{N}(X^{\omega}):\mathrm{R}^m\mapsto[0,1]^m$ – процедура нормализации обучающей выборки, X – нормализованная выборка; C_{FCM} – алгоритм кластеризации, формально представленный функцией $C_{FCM}(X;c,\ldots):\mathrm{R}^{n\times m}\mapsto M_{fcn}$, где c – количество элементов обучающей выбо-

рки, M_{fcn} – определено в (2), c – параметр алгоритма, задающей количество кластеров, а далее идут другие параметры FCM-алгоритма [6], [8]; v^* – критерий адекватности, например, (9) или (12), а $U^1 = \left(\underbrace{1,1,\ldots,1}_{n}\right)^T$ –единичный вектор, $\mathbf{G}^1 = \{\mathbf{g}_1\}$, где $\mathbf{g}_1$ вычисляется по формуле (7) при условии, что $U = U^1$. Диапазон значений $c = 2,3,\ldots,c_{max}$, причем $c_{max} \le \sqrt{n}$, где n – мощность множества X^{ω}.

В блоке построения модели выполняется анализ значений критерия качества. Для критерия (12) лучшим считается разбиение, для которого он минимален.

Таким образом, рассмотренная схема построения внутриклассовой структуры в виде модели FP^{ω} может быть применена для решения задачи распознавания образов с алфавитом классов $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$, где K – количество классов образов задачи, и известном множестве прецедентов Z. Сгруппируем элементы обучающей выборки Z по принадлежности классам образов, и представим эти группы как множества $X^{\omega_i} = \{(x_1, x_2, \ldots, x_m) \mid (x_1 x_2, \ldots, x_m, \omega_i) \in Z\}$, $\forall i = \overline{1, K}$.

В результате применения предложенного алгоритма для каждого класса образов получаем свою локальную модель FP^{ω_i}. А теперь выполним простую операцию объединения всех локальных моделей в единую нечеткую модель:

$$FM = < \left\{FP^{\omega_i}\right\}_{i=1}^{K}, D >, \tag{15}$$

где D – алгоритм принятия решения по данным модели. Далее определим D и покажем, как можно решать задачу распознавания образов в условиях неопределенности, используя только информацию данной модели (15).

3 Метод принятия решения

Для объекта $\mathbf{x} \in \mathrm{R}^m$ результатом работы алгоритма распознавания образов в условиях неопределенности в соответствии с рассмотренной постановкой задачи, является нечеткое множество $\tilde{\alpha} \in \tilde{\Omega}$, где $\tilde{\Omega}$ – пространство нечетких множеств над Ω,

$$\tilde{\alpha} = \{(\omega_1 \mid \mu_1), (\omega_2 \mid \mu_2), \ldots, (\omega_K \mid \mu_K)\}, \tag{16}$$

а μ_i – степень принадлежности $\mathbf{x}$ классу ω_i.

Основная идея предложенного в данной работе метода распознавания следующая. Настройка модели распознавания осуществляется по каждому классу образов отдельно, а на этапе принятия решения о том, к какому классу образов относится распознаваемый объект, используется информация по всем классам образов вместе. Преимуществом данного подхода является прозрачность полученной модели и следующая из способа ее построения способность к адаптации при появлении новых данных за счет дообучения.

Далее рассмотрим основные шаги алгоритма принятия решения $D(x^*; FM, \Lambda, \varepsilon) = \tilde{\alpha}$, где FM -модель – (15), $\tilde{\alpha}$ – в виде (16), $\Lambda = \left\{\left(p_{1f_i}, p_{2f_i}\right)\right\}_{i=1}^{m}$ – множество пар минимальных p_{1f_i} и максимальных p_{2f_i} значений, используемых при нормализации данных, f_i – информативный признак, $i = \overline{1, m}$, m – количество информативных признаков, ε –

пороговое значение; а x^* – новый объект, который необходимо распознать. При описании алгоритма множество центров кластеров, соответствующих модели FP^{ω_i}, и их количество, обозначим G^{ω_i} и c^{ω_i} соответственно.

Шаг 1. Выполним преобразование $\mathbf{x}^*$ в соответствии с Λ по формуле:

$$\mathbf{x} = \frac{\mathbf{x}^* - \mathbf{p}_1}{\mathbf{p}_2 - \mathbf{p}_1} \qquad (17)$$

где $\mathbf{p}_1$ и $\mathbf{p}_2$ как в формуле (8).

Шаг 2. Для всех $\omega_i, i = \overline{1, K}$ вычисляем μ_i, чтобы получить $\widetilde{\alpha}$ в виде (16). Переопределим $\omega = \omega_i, \mu = \mu_i$, $FP = FP^{\omega_i}$.

Вычислим расстояние от $\mathbf{x}$ до ближайшего центра кластера класса ω по модели FP:

$$l = \min_{\mathbf{g} \in \mathbf{G}^{\omega}} \left(d(\mathbf{x}, \mathbf{g})\right), \qquad (18)$$

где d определено в (4).

Определим множество B^{ω}, которое состоит из всех центров кластеров модели, кроме данных для класса ω:

$$B^{\omega} = \bigcup_{\substack{j=1 \\ j \neq i}}^{k} G^{\omega_k}, \qquad (19)$$

где i – индекс класса ω.

Для вычисления степени принадлежности μ определим следующую формулу:

$$\mu = \left(1 + \sum_{\mathbf{g} \in B^{\omega}} \frac{l^2}{d^2(\mathbf{x}, \mathbf{g})}\right)^{-1} \qquad (20)$$

где l вычисляется по формуле (18), B^{ω} определено в (19), а d определено в (4).

Посчитанные в (20) μ для каждого ω подставим в $\widetilde{\alpha}$ в виде (16).

Шаг 3. Выполним следующее преобразование для полученного на шаге 4 нечеткого множества $\widetilde{\alpha}$. Исключим из рассмотрения классы с малыми значениями степеней уверенности μ_i, заменив $\widetilde{\alpha}$ его α – срезом:

$$\mu_i = \begin{cases} 0, \text{если } \mu < \varepsilon \\ 1, \text{если } \mu \geq \varepsilon \end{cases}, \qquad (21)$$

где ε – параметр алгоритма распознавания, соответствующий пороговому значению.

Шаг 4. Чтобы оценить качество построенной модели введем следующую операцию приведения к четкости $H(\widetilde{\alpha}; \delta)$, которую будем называть операцией мультиклассового приведения к четкости с функцией отказа от распознавания:

$$H(\widetilde{\alpha}) = \begin{cases} (\omega_{l1}, \omega_{l2}, \ldots, \omega_{lk}), \text{если } \forall i, j, i \neq j \ (\mu_{l_i} - \mu_{l_j}) > \delta \\ \Delta, \text{иначе} \end{cases}, \qquad (22)$$

где $\delta \in [0,1]$ – параметр алгоритма, Δ – символ отказа от распознавания, l_1 – номер класса образов, степень принадлежности к которому максимальная $l_1 = \arg max_{\ j=1,c} \mu_j$, l_2 – номер класса образов, для которого степень принадлежности

максимальная среди оставшихся классов образов $l_2 = \arg max_{\ j=1,c \wedge j \neq l_1} \mu_j$, и так далее, $l_k = \arg max_{\ j=1,c \wedge j \neq l_1 \wedge j \neq l_2 \wedge \ldots \wedge j \neq l_{k-1}} \mu_j$, k – количество классов образов, к которым алгоритм относит рассматриваемый образец. Если $k > 1$, то образец $\mathbf{x}$ находится в области пересечения классов.

4 Результаты экспериментов

Предложенный метод распознавания применяется для решения задач идентификационной экспертизы в системах экспертного контроля. Под термином «идентификация» в данном случае не подразумевается общеизвестный математический термин, а только понятие установления тождественности между реальным объектом и существенными его характеристиками, заявленными в документации, в целях борьбы с фальсификациями. К таким задачам относится практическая задача идентификации производителя бензина [10], для которой проведен ряд экспериментов.

Для оценки качества алгоритма распознавания будем делить выборку прецедентов на обучающую и тестовую в соотношении 2:1 для каждого класса соответственно.

Для оценки качества алгоритма распознавания используется матрица неточностей Q. Элемент q_{ij} матрицы неточностей определяет количество элементов тестовой выборки, для которых правильным классом образов является класс ω_i, а алгоритмом распознавания он был отнесен к классу ω_j. В табл. 1 приведен пример матрицы неточности для задачи распознавания с двумя классами образов. Искусственно введенный класс ω_0 соответствует отказу от распознавания, как в функции (22).

Таблица 1 – Матрица неточностей

Метод / Модель	ω_0	ω_1	ω_2
ω_1	q_{10}	q_{11}	q_{12}
ω_2	q_{20}	q_{21}	q_{22}

Пусть p_i, $i=\overline{1,K}$ – априорная вероятность встречаемости элементов данного класса для задачи распознавания, $\sum_{i=1}^{K} p_i = 1$. Вычислим вероятность событий, при которой алгоритм определил, что объект класса ω_i является объектом класса ω_j, $i=\overline{1,K}, j=\overline{0,K}$:

$$p_{ij} = \frac{q_{ij}}{\sum_{k=0}^{K} q_{ik}}, \tag{23}$$

где q_{ij} – элементы матрицы неточности.

Рассмотрим два типа ошибки для алгоритма с функцией отказа. Первый тип ошибки будет учитывать только ситуацию, когда алгоритм дал неправильный ответ, и будет определять вероятность ошибки алгоритма:

$$A^I = \sum_{i=1}^{K} p_i \left(\sum_{\substack{j=1 \\ j \neq i}}^{K} p_{ij} \right). \tag{24}$$

Вероятность отказов рассчитаем по формуле:

$$R = \sum_{i=1}^{K} q_{i0} . \tag{25}$$

Тогда ошибка второго типа будет включать как неправильный ответ, так и отказ от распознавания и ее вероятность вычислим как:

$$A^{II} = A^{I} + R \tag{26}$$

Далее в результате экспериментов качество модели оценивается по формулам (24), (25) и (26).

Задача классификации вин. Первая группа экспериментов была выполнена для бенчмарк-задачи классификации вин, приведенной в [11]. Для каждой пробы вина известны 13 его характеристик, перечислим некоторые из них: содержание спирта, яблочной кислоты, флавоноидов и т.д. Даны результаты экспериментов для 178 проб вина для трех типов вин, для которых известно, из какого сорта винограда они произведены. Необходимо построить модель классификации, по которой можно было бы определить для неизвестной пробы вина сорт винограда, из которого оно произведено.

Алфавит классов образов для данной задачи распознавания состоит из трех элементов $\Omega = \{V1, V2, V3\}$, количество информативных признаков $m = 13$.

В табл. 2 приведены значения критериев (12) v_{XB} и (9) v_{PC}.

Таблица 2 – Значение критериев адекватности для задачи классификации вин

c	v_{XB}, V1	v_{XB}, V1	v_{XB}, V1	v_{PC}, V1	v_{PC}, V2	v_{PC}, V3
1	–	–	–	1	1	1
2	**11.47**	**27.88**	**8.75**	0.51	0.50	0.52
3	59.12	124.25	43.36	0.34	0.34	0.36
4	308.73	438.12	104.95	0.25	0.25	0.27

Результат работы алгоритма при $\delta = 0.01$ $A^{I} = A^{II} = 5{,}79\%$ на тестовой выборке и $A^{I} = 0.69\%$, $A^{II} = 1.52\%$ на обучающей выборке в предположении, что классы встречаются с равной априорной вероятностью $p_{V1} = p_{V2} = p_{V3} = \frac{1}{3}$.

Результаты других исследовании для задачи классификации вин находятся в пределах 0% – 10% и указывают на то, что классы образов сильно пересекаются [11]. Таким образом, полученный результат подтверждает эффективность предложенного алгоритма.

Выполним анализ значений коэффициентов адекватности, приведенных в табл. 2. Минимальное значение v_{XB} для всех трех классов достигается при $c = 2$, а значение критерия $v_{PC} \approx \frac{1}{c}$, что соответствует «предельно нечеткому разбиению», $\overline{U}$, что было рассмотрено при описании свойств во второй части статьи. Таким образом, можно сделать вывод, что классы образов для данной задачи не имеют внутренней структуры, $c = 1$.

Также значение v_{XB} для V2 при $c = 2$ больше, в сравнении со значением данного критерия для классов V1 и V3, что соответствует тому факту, что алгоритм распознавания ошибался именно на представителях класса V2.

Задача идентификации производителя бензина. Данная практическая задача возникает в лабораториях контроля качества нефтепродуктов при их реализации и направлена на борьбу с фальсификациями топлива. В качестве классов образов выступает бензин определенного производителя и определенной марки, а информативными призна-

ками являются физико-химические показатели образца. Предложенный метод распознавания используется в автоматизированной системе идентификации нефтепродуктов, разработанной для компании ООО «Параллель-М ЛТД».

Значение ошибки первого типа для обучающих выборок за разные периоды не выше 4%, а ошибки второго типа не выше 12%. Такой показатель обусловлен тем, что классы сильно пересекаются.

Выводы

В данной статье предложен новый алгоритм распознавания для решения задачи в условиях неопределенности, который может быть использован при разработке открытых систем распознавания. Представление результата работы алгоритма распознавания в виде нечеткого множества, описывающего степень похожести объекта на каждый класс образов, позволяет решить проблему неопределенности.

Предложенный алгоритм разработан и апробирован на задачах идентификационной экспертизы. Для этих задач актуальны основные преимущества метода: прозрачность модели и возможность расширять модели при появлении новых данных.

Литература

1. Воронцов К.В. Комбинаторный подход к оценке качества обучаемых алгоритмов / К.В. Воронцов // Математические вопросы кибернетики. – 2004. – № 13. – М. : Физматлит, 2004. – С. 5-36.
2. Kuncheva L.I. Fuzzy Classifier Design / Kincheva L.I. – Springer Verlag, 2000. – 315 p.
3. Ishibuchi H. Classification and Modeling with Linguistic Information Granules / H. Ishibuchi, T. Nakashima, M. Nii. – Spriger, 2005. – 307 p.
4. Сергиенко Р.Б. Коэволюционный генетический алгоритм решения сложных задач условной оптимизации / Р.Б. Сергиенко, Е.С. Семенкин // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнёва. – № 2 (23). – 2009. – С. 17-21.
5. Chen N. Fuzzy Classification Using Self-Organizing Map and Learning Vector Quantization / N. Chen // CASDMKM 2004. Lecture Notes in Artificial Intelligence. – Springer-Verlag, 2004. – V. 3327. – P. 41-50.
6. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing / [J.C. Bezdek, J. Keller, R. Krisnapuram, R. Pal]. – New York : Springer Verlag, 2005. – 785 p.
7. Handbook of Granular Computing / W. Pedrycz, A. Skowron, V. Kreinovich, [eds.]. – N.-Y. : John Wiley & Sons,2007. – 1116 p.
8. Pal N. On Cluster Validity for the Fuzzy c-Means Model / N. Pal, J. Bezdek // IEEE Transactions on Fuzzy System. – 1995. – V. 3, № 3. – P. 370-379.
9. Maksimova A. The Model of Data Presentation with Fuzzy Portraits for Pattern Recognition / A. Maksimova // International Journal of Computing. – 2012. – Vol. 11, Issue 1. – P. 17-24.
10. Козловский В.А. Нечеткая система распознавания образов для решения задач классификации жидких нефтепродуктов / В.А. Козловский, А.Ю. Максимова // Наукові праці ДонНТУ, Серія «Інформатика, кібернетика та обчислювальна техніка». – 2011. – №13 (185). – С. 200-205.
11. Frank A. UCI Machine Learning Repository. Wine Data Set. / A. Frank, Asuncion Irvine C.A. [Электронный ресур]. – University of California, School of Information and Computer Science. – 2010. – Режим доступа : [http://archive.ics.uci.edu/ml/ datasets/Wine].

Literatura

1. Vorontsov K.V. Combinatorial approach to evaluating the quality of training algorithms / K.V. Vorontsov // Mathematical Problems of Cybernetics. - 2004. - № 13 - Moscow: Fizmatlit, 2004. – P. 5-36.
2. Kuncheva L. I. Fuzzy Classifier Design / Kincheva L.I. – Springer Verlag, 2000. – 315 p.
3. Ishibuchi H. Classification and Modeling with Linguistic Information Granules / H. Ishibuchi, T. Nakashima, M. Nii. – Spriger, 2005. – 307 p.

4. Sergienko R.B. Co-evolutionary genetic algorithm to solve complex constrained optimization problems / R.B. Sergienko, E.S. Semenkin // Bulletin of the Siberian State Aerospace University named after Academician M.F. Reshetnev. – № 2 (23). – 2009. – P. 17-21.
5. Chen N. Fuzzy Classification Using Self-Organizing Map and Learning Vector Quantization / N.Chen // CASDMKM 2004. Lecture Notes in Artificial Intelligence. – V. 3327. – Springer-Verlag, 2004. – P. 41-50.
6. Bezdek J.C. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing / J.C. Bezdek, J. Keller, R. Krisnapuram, R. Pal. – New York : Springer Verlag, 2005. – 785 p.
7. Handbook of Granular Computing / W. Pedrycz, A. Skowron, V. Kreinovich, [eds.]. – N.-Y. : John Wiley & Sons,2007. – 1116 p.
8. Pal N. On Cluster Validity for the Fuzzy c-Means Model / N. Pal, J. Bezdek // IEEE Transactions on Fuzzy System. – 1995. – V. 3, № 3. – P. 370-379.
9. Maksimova A. The Model of Data Presentation with Fuzzy Portraits for Pattern Recognition / A. Maksimova // International Journal of Computing. – 2012. – Vol. 11, Issue 1. – P. 17-24.
10. Kozlowskyj V.A. Fuzzy pattern recognition system to solve the problems of liquid petrolium classification / V.A. Kozlowskyj, A. Yu. Maksymova // Proceedings of Donetsk National Technical University, Series «Informatics, Cybernetics and Computer Science». – 2011. – № 13 (185). – S. 200-205.
11. Frank. A. UCI Machine Learning Repository. Wine Data Set / A. Frank, Asuncion Irvine C.A. – University of California, School of Information and Computer Science. 2010. – [http://archive.ics.uci.edu/ml/ datasets/Wine].

RESUME

A. Maksymova

Decision Making Method for Classifying Models Based on Intra-class Clustering on FCM-algorithm

This paper deals with pattern recognition method in conditions of uncertainty for open recognition systems. Decision support systems for identification examination are example of such systems, we discuss here.

The formally formulation of the problem is presented. An algorithm for intra-class clustering based on FCM-algorithm is worked out. The Xie-Beni index is used to solve problem of searching optimal number of clusters.

An algorithm for decision making is worked out. This algorithm calculates the membership degrees of recognizable object using the distances to prototypes of clusters. The function of multiclass defuzzyfication with rejection option is introduced, that allows to take into account the uncertainty of the result of the decision.

The experiments confirmed the effectiveness of the method on reference problem of wine classification and on the petroleum identification problem. The work results are used in automated system for liquid petroleum checking.

Статья поступила в редакцию 10.06.2013.